Lecture 14:

Udlearn. 17
2012

Learning halfspaces (linear threshold fctns)

Def. $h(x) = \text{sign}(w \cdot x - \theta)$ is a "halfspace function"

$\text{sign}(x) \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{o.w} \end{cases}$

Thm Let h be a halfspace over $\{\pm 1\}^n$
then h has Fourier concentration $\alpha(\varepsilon) = \frac{c}{\varepsilon^2}$

$$\left(\text{i.e. } \sum_{|S| \geq \frac{c}{\varepsilon^2}} \hat{h}(S)^2 \leq \varepsilon\right)$$

(Will prove soon)

Corr low degree algorithm learns halfspaces under uniform distribution with $n^{O(1/\varepsilon^2)}$ uniform samples.

(Actually can learn in $O(n^5)$ but we'll get a "big win" from this approach soon...)

Key idea: Noise sensitivity

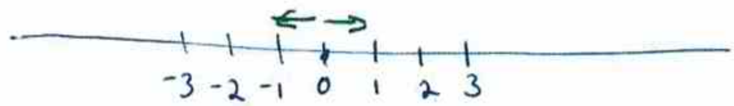
3) $f(x) = Maj(x_1 \cdots x_n)$

$ns_\varepsilon(f) = O(\sqrt{\varepsilon})$

<u>Sketch</u>

- $Maj(x) \sim$ random walk on line $\boxed{\text{starting at } 0}$

$Pr(L) = \frac{1}{2}$ $Pr(R) = \frac{1}{2}$

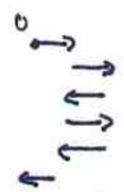


$x_i = 1 \Rightarrow R \times 2$

$x_i = -1 \Rightarrow L \times 2$

$\sum x_i =$ endpt of walk (if start at 0)

e.g. $x = (1\,1\,{-1}\,1\,{-1}\,{-1})$



ends at 0

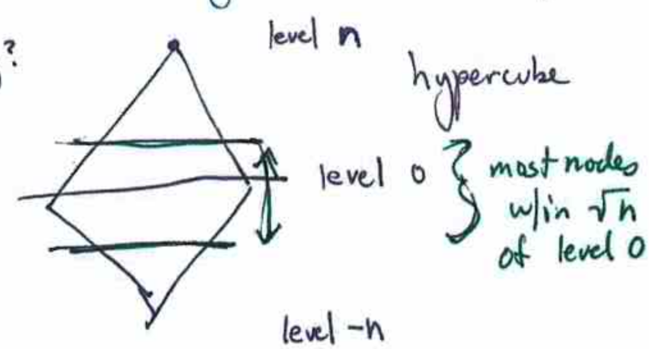
equivalent process: pick random pt on hypercube & output the level

Well known fact:

expected distance from startpt ($|x_1 + x_2 + \ldots + x_n|$) after n steps is $\sqrt{n}$

& likely to be close to expectation

why?



- $N_\varepsilon(x) \sim$ random walk on εn $\underbrace{\text{bits}}_{\text{each}}$ has twice value taken

since change $+1 \to -1$, $-1 \to +1$

expected displacement $2\sqrt{\varepsilon n}$

Another view:
- take walk according to x
- then take walk according to $N_\varepsilon(x)$

6) any f:

Thm $f: \{\pm 1\}^n \to \{\pm 1\}$

$$NS_\varepsilon(f) = \frac{1}{2} - \frac{1}{2} \sum_S (1-2\varepsilon)^{|S|} \hat{f}(S)^2$$

note for parity fctns, this gives

$$NS_\varepsilon(\chi_S) = \frac{1}{2} - \frac{1}{2}\cdot(1-2\varepsilon)^{|S|} \quad ✓$$

Pf.

$$NS_\varepsilon(f) = \Pr_{\substack{x \\ y \in N_\varepsilon(x)}} [f(x) \neq f(y)]$$

$$= E_{\substack{x, \\ y \in N_\varepsilon(x)}} [\mathbb{1}_{f(x) \neq f(y)}]$$

$$= E_{xy}\left[\frac{(f(x)-f(y))^2}{4}\right] = \frac{1}{4} E\,[\underbrace{f(x)^2 + f(y)^2}_{\text{always } 1} - 2 f(x) f(y)]$$

$$= \frac{1}{2} - \frac{1}{2} E[f(x) f(y)]$$

$$= \frac{1}{2} - \frac{1}{2} E\left[\sum_S \hat{f}(S) \chi_S(x) \sum_T \hat{f}(T) \chi_T(y)\right]$$

$$= \frac{1}{2} - \frac{1}{2} \sum_{S,T} \hat{f}(S) \hat{f}(T)\, E[\underbrace{\chi_S(x) \chi_T(y)}_{=0 \text{ if } S \neq T}]$$

} show using standard techniques using pairing on $i \in S \Delta T$

but, if $S = T$?

$$E_{xy}[\chi_S(x)\chi_T(y)] = 1 \cdot \Pr[\chi_S(x) = \chi_S(y)]$$

$$+ (-1) \Pr[\chi_S(x) \neq \chi_S(y)]$$

$$= 1 - 2 \underbrace{\Pr[\chi_S(x) \neq \chi_S(y)]}_{NS_\varepsilon(\chi_S) = \frac{1}{2} - \frac{1}{2}(1-2\varepsilon)^{|S|}}$$

$$= (1-2\varepsilon)^{|S|}$$

$$= \frac{1}{2} - \frac{1}{2} \sum_S (1-2\varepsilon)^{|S|} \hat{f}(S)^2$$

∎

Back to ½ spaces:

<u>Corr 2</u> for ½ space $h: \{\pm1\}^n \to \{\pm1\}$

$$\sum_{|s| \geq O(\frac{1}{\varepsilon^2})} \hat{f}(s)^2 \leq \varepsilon$$

<u>Pf</u>

for ½ space $ns_\varepsilon(h) \leq 8.8\sqrt{\varepsilon}$ so use $\beta(\varepsilon) = 8.8\sqrt{\varepsilon}$

$\beta^{-1}(x) = \left(\frac{x}{8.8}\right)^2$

$\beta^{-1}\left(\frac{\varepsilon}{2.32}\right) \leq \left(\frac{\varepsilon}{\underbrace{2.32 \times 8.8}_{(20.416)}}\right)^2 \leq \left(\frac{\varepsilon}{20}\right)^2$

so for $m \geq \frac{1}{\left(\frac{\varepsilon}{20}\right)^2}$

□

⇒ Can learn any ½ space over uniform with $n^{O(1/\varepsilon^2)}$ random examples

but, can do <u>a lot</u> better!

still, technique extends ...

Learn any fctn of K ½ spaces:

$h_1 \dots h_K$ are K half spaces

$g: \{\pm1\}^K \to \{\pm1\}$ Boolean fctn

$f: \{\pm1\}^n \to \{\pm1\}$ st $f(x) = g(h_1(x), \dots, h_K(x))$

<u>Thm</u> $ns_\varepsilon(f) = 8.8K\sqrt{\varepsilon}$